Theme: Examining if there are trends in allegations/complaints between identity groups of civilians, identity groups of officers, and relationships between identity groups of officers and civilians.

*An "identity group" is a group of people that have the same race, gender, age. We will need all three of these factors to determine a person's identity group.*

Description and summarization:

1. Which identity groups filed the most complaints?
2. What percentage of unique officers have multiple allegations of any kind against them? (from "data allegations" in cpdb)
3. Which identity groups are most often victims?
4. What percentage of unique officers have multiple victim allegations against them? (from "data allegations" in cpdb)

Data Integration:

1. Of the officers named in settlements, are there trends in the identity groups of the officers? (using info from settlement data and cpdb)
2. Of the officers named in settlements, are there trends in the identity groups of the respective plaintiffs? (using info from settlement data and cpdb)
3. Of the officers named in settlements, are there trends in the relationship of identity groups between the officers and plaintiffs? (using info from settlement data and cpdb)
4. Do police officers of certain identity groups have more allegations against them in low income areas than in higher income areas? (using info from economic development data and cpdb)

Workflow analytics:

1. Are certain types of allegations made against officers rising or falling faster compared to the rate of change of total allegations made?
2. Are the percentages of allegations made by civilians of any identity group rising or falling compared to the rates of total allegations *and* to the rates of population demographics in Chicago?
3. Are the percentages of allegations against officers of any identity group rising or falling compared to the rates of total allegations *and* to the rates of population demographics in Chicago?

Machine learning:

1. Given an officer's identity group and a complaint report details, can we predict the identity group of the victim? (features: victim_gender, victim_race, officer_age, officer_race, officer_gender)
2. Given data from years 1-10, can we predict how the rate of complaints (by white women civilians against black men officers) that result in action will change over time for years 11-15? What about for complaints by black women civilians against black men officers? (features: victim_gender, victim_race, officer_race, officer_gender, final_outcome, investigators)

3. Given a complainant’s identity group and the accused officer's identity group, can we predict whether the complaint will result in action? (features: victim_gender, victim_race, officer_age, officer_race, officer_gender, final_outcome, investigators)

Modeling with Neural Networks:

1. Predict the identity group of an officer given a TRR or complaint report (or the identity group of the victim if applicable) by training the model on the text of complaints and TRRs from different identity groups.

Visualization:

1. Which pairings between identity groups of officers and their respective accusers are most common? (One axis is officer identity group, the other axis is accuser identity group. Each intersection of officer and accuser is a datapoint of number of instances. Data points are larger with greater number of instances.)
2. Which identity groups make an officer most likely to have an allegation? (Pie chart/bar graph)
3. What identity groups of civilians make them more likely to file complaints/accusations for each type of harassment? (Pie chart/bar graph)
4. What identity groups of officers make them more likely to have complaints/accusations for each type of harassment filed against them? (highlight table)

data_allegation

| id | crid | summary | add2 | city | incident_date | point | source | beat_id | is_officer_complaint | add1 | location | old_complaint_address |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 65308 | 1012597 | | S AVERS AVE | CHICAGO IL 60623 | 2007-12-03 18:00:00-06 | 0101000020E6100000B168DF8126EE55C0CC1022742BEB4440 | | 159 | f | 30XX | Public Way - Other | |
| 65309 | 1012598 | | W THOMAS ST | CHICAGO IL 60651 | 2007-12-11 18:00:00-06 | 0101000020E6100000BD5301F73CEF55C0721CD36E4FF34440 | | 66 | f | 44XX | Public Way - Other | |
| 65310 | 1012601 | | S EMERALD AVE | CHICAGO IL 60609 | 2007-12-08 18:00:00-06 | 0101000020E6100000731992EE42E955C0D1ABA690D0E74440 | | 112 | f | 45XX | Public Way - Other | |
| 65311 | 1012602 | | S MOZART ST | CHICAGO IL | 2007-12-09 18:00:00-06 | 0101000020E610000083C53C86A2EC55C006A051BAF4EE4440 | | 137 | f | 11XX | Public Way - Other | |
| 65312 | 1012603 | | S INDIANA AVE | CHICAGO IL 60628 | 2007-12-11 18:00:00-06 | 0101000020E6100000E8C072848CE755C0F57700D589D74440 | | 217 | t | 115XX | Public Way - Other | |

data_officerallegation

 id  | start_date |  end_date  | officer_age | recc_finding | recc_outcome | final_finding | final_outcome       | final_outcome_class | allegation_id | allegation_category_id | officer_id | disciplined

-----+------------+------------+-------------+--------------+--------------+---------------+----------------------+---------------------+---------------+------------------------+------------+-------------

 239 | 1967-10-21 | 1980-08-01 |             |              | Unknown      | SU            | 30 Day Suspension    |                     |        111233 |                    207 |      16796 | t

 240 | 1970-08-13 | 1973-09-15 |             |              | Unknown      | SU            | Suspended Over 30 Days |                     |        111234 |                    132 |      21822 | t

 241 | 1972-02-18 | 1973-02-02 |             |              | Unknown      | SU            | Suspended Over 30 Days |                     |        111235 |                    201 |        214 | t

 242 | 1970-09-22 | 1971-10-15 |             |              | Unknown      | SU            | Suspended Over 30 Days |                     |        111236 |                     49 |      11652 | t

 243 | 1970-10-19 | 1987-03-09 |             |              | Unknown      | SU            | Suspended Over 30 Days |                     |        111237 |                    132 |       3089 | t

data_complainant

 id  | gender | race  | age | allegation_id | birth_year

-----+--------+-------+-----+---------------+------------

   1 | M      | White |  50 |         57419 |

  12 | M      | White |     |         58144 |

 161 | F      | Black |  39 |         69105 |

 162 | M      | White |  55 |         69379 |

 326 | M      | White |  67 |         79867 |

data_victim

  id  | gender | race  | age | allegation_id | birth_year

------+--------+-------+-----+---------------+------------

 1863 | M      | Black |     |         64378 |

 2199 | F      | Black |     |         56991 |

 3075 | M      | Black |     |         57910 |

 3590 | M      | White |     |         57236 |

 3759 | M      | Black |     |         56943 |       1969